

25 January 1967

Material Test Procedure 3-1-002*
Aberdeen Proving Ground

U. S. ARMY TEST AND EVALUATION COMMAND
BACKGROUND DOCUMENT

CONFIDENCE INTERVALS AND SAMPLE SIZE

D D C
RECEIVED
FEB 19 1971
B

1. INTRODUCTION

The confidence interval is a useful tool in the analysis and presentation of test results. It is an estimate of a range of values that is stated to contain the value of some numerical characteristic of a statistical population, the interval in any given instance being computed from a specific sample taken at random from the population. In some cases the confidence interval is specified and sample size determined. The probability that the statement is true is the "confidence coefficient" associated with the given confidence interval. This is the equivalent to saying that a given percent confidence interval places a certain value between two fixed "confidence limits".

— / Many of the misunderstandings and misuses of confidence intervals can be avoided if the following properties are kept in mind: (1) the interval is a statement about a characteristic of a statistical population, (2) the interval is derived from a random sample of that population and is subject to sampling variation, i.e., another sample and a repetition of the test would be expected to yield a different interval; and (3) the confidence coefficient is not associated with a particular interval but with the particular method of calculating it. ()

2.0 CONCEPTS

Population characteristics such as the mean and standard deviation can rarely be determined exactly because of the unavailability of all the individuals in the population, the expense of examining every individual, or the destructive nature of the examination. Consequently population characteristics must be inferred from an examination of a small part of the population - a randomly selected sample. The general character of the population often is known or can safely be assumed to be of a type that can be described by a known mathematical expression (distribution function) containing one or more unknown constants (parameters). When that is the case, estimation of properties of the population is reduced to a problem of estimation of the unknown parameters of the distribution functions. The estimates are obtained by selecting a sample, performing the experiment or test, making the required observations (e.g., measuring a dimension or counting the number of individuals in one of two categories), and making the calculations implied by an appropriate formula. Thus a single-valued estimate (usually called a point estimate) of the parameter is obtained. Experience verifies that a repetition of those steps for the same population almost always yields a different result and a different estimate of the same parameter. Because of this statistical variation exhibited by single-valued estimates, an interval estimate that could take into account that variation would be preferred in certain applications.

The derivation of a useful interval estimate from a single sample is possible. The procedure makes use of probability theory which, in turn,

* SUPERSEDES INTERIM PAMPHLET 90-31

-1-

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
Springfield, Va. 22151

DISTRIBUTION STATEMENT 4
Approved for public release;
Distribution Unlimited

AD 718229

supposes that some knowledge of the distribution of the population is available. For example, if the population is normally distributed and a random sample is available, confidence intervals for the mean and for the standard deviation can be calculated. Although the confidence interval is a statement that the parameter lies within the interval, it is not an absolute inequality. An essential part of the statement is the confidence coefficient. (The confidence interval is meaningless without the confidence coefficient.) Since the interval is derived from a sample, repeated sampling and calculation would be expected to yield different intervals. Some of them will bracket the parameter and some will not. The confidence coefficient is simply the proportion of such intervals - assuming the entire sampling, calculating, etc., process were repeated a large number of times - that may be expected to bracket the population parameter. A particular interval, however, either brackets the parameter or it does not. The confidence coefficient is descriptive of the entire process rather than of a single interval.

Prior to the conduct of sampling and experimentation it is appropriate to consider the probability that an interval calculated from the outcome of the experiment will bracket the population parameter. That probability is numerically equal to the confidence coefficient. After the observations are obtained, however, the chance elements are no longer present and the concept of probability is no longer applicable. The confidence coefficient associated with a particular interval must, therefore, not be interpreted as a probability.

2.1 STEPS

Procedures for calculating confidence intervals are described in a through d below. Paragraphs 3.1 through 3.1.6 give step-by-step examples of procedures for determining confidence intervals in seven common situations. In every instance it is assumed that the observations from a random sample are available. The number of observations is denoted by n , the observations are called x_i (the subscript i may be omitted when no ambiguity results), and the confidence coefficient is γ .

a. The first step is to decide on the appropriate type of confidence interval. A confidence interval may be "two sided" or "one sided". A two-sided interval states that the parameter lies between two specific values inside the possible range of the parameter. If the interval states merely that the parameter is greater than some specific value with no statement about an upper limit, it is called a lower confidence limit. This would be a one-sided interval. An upper confidence limit is similarly defined. The graphical aids in the Appendix (with one exception) are made specifically for two-sided intervals. They can, however, be used for calculation of one-sided intervals provided a simple adjustment of the confidence coefficient is made.

b. The confidence coefficient is selected. Typical coefficients in common use are 0.90, 0.95, and 0.99. In general, the larger the confidence coefficient, the wider the interval will be. Since, in a sense, the coefficient is a measure of the "strength" of the statement, that relationship is clearly consistent. The graphs in Figures 1 through 4 show curves for confidence coefficients of 0.80, 0.90, 0.95, and 0.99 for two-sided intervals. These same curves are for 0.90, 0.95, 0.975, and 0.995 (respectively) for one-sided intervals.

N O T I C E

**THIS DOCUMENT HAS BEEN REPRODUCED FROM THE
BEST COPY FURNISHED US BY THE SPONSORING
AGENCY. ALTHOUGH IT IS RECOGNIZED THAT CER-
TAIN PORTIONS ARE ILLEGIBLE, IT IS BEING RE-
LEASED IN THE INTEREST OF MAKING AVAILABLE
AS MUCH INFORMATION AS POSSIBLE.**

c. The appropriate table is entered to obtain one or more factors. Each table is designed for ease of use. Each has been derived from an accepted statistical table, such as the Student-t, Chi Square, standard normal, etc.

d. The formulas incorporating the observations and the factors from c are calculated.

2.2 USES

Confidence intervals are most frequently used in the analysis of test results, the application illustrated in this pamphlet. They also may be used in the planning of experiments, as an aid in determining the number of samples needed to produce a desired interval with a specified confidence coefficient. Dependence of the interval on the test results greatly limits this application. For a few intervals, however, the dependence is not a limitation. For example in paragraph 3.1 the width ($2K\sigma_0$) of the two-sided interval for the mean does not depend on the observations. If an interval of a specified width is desired, various combinations of γ and n can be found to produce that width. (The limits of the graph will restrict the number of combinations.) Similarly, in paragraph 3.1.3 the ratio of the upper to the lower limits of the two-sided interval for the standard deviation is independent of the observations and depends only on γ and n . For the majority of experiments, however, the planner is obliged to employ assumed results in order to use the confidence interval as an aid in determining sample size.

As a tool for analyzing data generated in planned experiments, the confidence interval appears simple and objective. There is thus a growing tendency to rely heavily on it to the exclusion of other valid methods of analysis. A simple statistical technique can rarely take into account all the information relevant to the evaluation of the performance of a complex item. In the testing of materiel one often wishes to obtain high assurance of very low (or very high) incidence of bad (or good) performance. An example is the desire for high assurance of nearly no premature functioning of high-explosive ammunition. If a firing test is made, the confidence interval for the expected premature rate derived by the methods of paragraph 3.1.6 is always disturbingly large even when no premature firing occurs. The use of this statistical technique alone will not provide the desired result. One must recognize that such tests can only support or refute an expectation of low incidence of premature firing based on previous experience and engineering analyses. In essence, one is applying an "a priori" assumption concerning the characteristics of the test item - an assumption not taken into account by the statistical technique. A discussion of this limitation in the use of statistical methods is contained in Reference 4.

3.0 EXAMPLES

3.1 Mean of a Normal Population, Standard Deviation Known

From prior experiments or for theoretical reasons the standard deviation of the population is known. An experiment is conducted to determine the mean of the population.

μ = unknown population mean

σ_0 = known standard deviation

n = number of observations

x_i ($i = 1, 2, \dots, n$) = the observations

$\bar{X} = \frac{1}{n} \sum x_i$, the sample mean

a. The type of confidence interval appropriate for the desired statement is determined:

- 1) A two-sided interval if it is desired to place μ between two finite limits.
- 2) A one-sided interval if it is desired to place a lower or upper limit on μ .

The confidence coefficient applies to any one of the three possible intervals but not to two or more simultaneously.

b. The desired confidence coefficient, γ , is selected. Values of γ available from the graphs are 0.80, 0.90, 0.95, and 0.99 for two-sided intervals and 0.90, 0.95, 0.975, and 0.995 for one-sided intervals.

c. The curves of Figure 1 are used to obtain a factor K (ordinate) by selecting the appropriate curve and entering the graph with n as the abscissa. Curves are labeled with the confidence coefficients for two-sided intervals. For a one-sided interval the curve is selected as follows: If γ' is the desired confidence coefficient, use the curve for $\gamma = 2\gamma' - 1$.

d. Confidence intervals are calculated by substitution in the appropriate formula:

$$\bar{X} - K\sigma_0 \leq \mu \leq \bar{X} + K\sigma_0 \quad \text{two-sided interval}$$

$$\bar{X} - K\sigma_0 \leq \mu \quad \text{one-sided interval (lower limit)}$$

$$\mu \leq \bar{X} + K\sigma_0 \quad \text{one-sided interval (upper limit)}$$

Illustration: Muzzle velocities of 8 rounds of projectile, 105-mm, HE, M1 selected at random from one lot of ammunition (Lot A) were measured to estimate the lot average. Records of firings of other lots from this production show that the standard deviation of velocity is 4 fps. A two-sided 95 percent confidence interval for the mean of Lot A is desired.

μ = unknown mean of Lot A, fps

σ_0 = 4 fps

n = 8

The observed muzzle velocities are as follows:

<u>Round No.</u>	<u>Muzzle Velocity, fps</u>
1	1551
2	1546
3	1554
4	1556
5	1549
6	1546
7	1547
8	1548

$$\bar{X} = \frac{12397}{8} = 1549.6$$

Reference to the curve for $\gamma = 0.95$ in Figure 1 gives $K = 0.69$ when $n = 8$. The confidence interval is obtained by substitution in the formula in d.

$$1549.6 - 0.69 (4) \leq \mu \leq 1549.6 + 0.69 (4)$$

$$1546.8 \leq \mu \leq 1552.4$$

It is asserted with 95 percent confidence that the average muzzle velocity of rounds of Lot A is between 1546.8 and 1552.4 fps.

3.1.1 Mean of A Normal Population, Standard Deviation Unknown

The experiment is conducted to determine the mean of the population. Data from the experiment are used to derive an estimate of the population standard deviation, i.e., no prior knowledge of the standard deviation is required.

μ = unknown population mean

n = number of observations

x_i ($i = 1, 2, \dots, n$) = the observations

$$\bar{X} = \frac{1}{n} \sum X_i, \text{ the sample mean}$$

$$S = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}, \text{ the sample standard deviation}$$

$$\text{or } S = \sqrt{\frac{n \sum X^2 - (\sum X)^2}{n(n - 1)}}, \text{ computing formula}$$

a. The type of confidence interval appropriate for the desired statement is determined:

- 1) A two-sided interval if it is desired to place μ between two finite limits
- 2) A one-sided interval if it is desired to place a lower or upper limit on μ

The confidence coefficient applies to any one of the three possible intervals but not to two or more simultaneously.

b. The desired confidence coefficient, γ , is selected. Values of γ available from the graphs are 0.80, 0.90, 0.95, and 0.99 for two-sided intervals and 0.90, 0.95, 0.975, and 0.995 for one-sided intervals.

c. The curves of Figure 2 are used to obtain a factor K (ordinate) by selecting the appropriate curve and entering the graph with n as the abscissa. Curves are labeled with the confidence coefficients for two-sided intervals. For a one-sided interval the curve is selected as follows: If γ' is the desired confidence coefficient, use the curve for $\gamma = 2\gamma' - 1$.

d. Confidence intervals are calculated by substitution in the appropriate formula:

$$\bar{X} - KS \leq \mu \leq \bar{X} + KS \text{ two-sided interval}$$

$$\bar{X} - KS \leq \mu \quad \text{one-sided interval (lower limit)}$$

$$\mu \leq \bar{X} + KS \text{ one-sided interval (upper limit)}$$

Illustration: An aircraft attempts to fly over and locate a ground station by following a signal emitted from the station. When the aircraft crew determines that the craft is over the station, a signal is sent and the actual location of the aircraft at that instant is determined by ground instrumentation. This experiment was performed six times under one set of conditions. It is desired to estimate the average horizontal distance between the aircraft and the ground station when the crew determined the craft to be over the station. A one-sided (upper), 95 percent confidence interval for the average distance is desired. No prior knowledge of the standard deviation is available.

μ = unknown population mean of horizontal distance
between aircraft and ground station

$n = 6$

The observations are as follows:

<u>Run No.</u>	<u>Horizontal Distance, m</u>
1	50
2	73
3	43
4	54
5	60
6	37

$$\bar{X} = \frac{317}{6} = 52.8$$

$$s^2 = \frac{n \sum X^2 - (\sum X)^2}{n(n-1)} = \frac{6(17563) - (317)^2}{30} = 162.97$$

$$S = 12.77$$

The appropriate curve in Figure 2 is selected. Since a one-sided interval with 95 percent confidence is required, the curve is taken as $2(0.95) - 1 = 0.90$, i.e., the curve $\gamma = 0.90$ is used. For $n = 6$, $K = 0.82$.

The confidence interval for μ is

$$\mu \leq \bar{X} + KS$$

$$\mu \leq 52.8 + 0.82(12.77)$$

$$\mu \leq 63.3$$

It is asserted with 95 percent confidence that the average horizontal distance between aircraft and ground station, when the aircraft crew determines that the craft is over the station, is no greater than 63.3 meters.

3.1.2 The Standard Deviation Of A Normal Population, Mean Known

From prior experiments or for theoretical reasons the mean of the population is known. An experiment is conducted to estimate the standard deviation of the population.

μ_0 = known population mean

σ = unknown standard deviation

n = number of observations

X_i ($i = 1, 2, \dots, n$) = the observations

$$S = \sqrt{\frac{\sum (X_i - \mu_0)^2}{n}}, \text{ the sample standard deviation}$$

$$S = \sqrt{\frac{\sum X_i^2}{n} - \mu_0^2}, \text{ computing formula}$$

a. The type of confidence interval appropriate for the desired statement is determined:

- 1) A two-sided interval if it is desired to place σ between two positive, finite limits.

- 2) A one-sided interval if it is desired to place a lower positive limit on σ , or to place an upper limit on σ .

The confidence coefficient applies to any one of the three possible intervals but not to two or more simultaneously.

b. The desired confidence coefficient, γ , is selected. Values of γ available from the graphs are 0.80, 0.90, 0.95, and 0.99 for two-sided intervals and 0.90, 0.95, 0.975, and 0.995 for one-sided intervals.

c. Figure 3 contains a pair of curves for each value of γ . If a two-sided interval is desired, a value K_1 is read from the lower curve of the pair and K_2 is read from the upper curve. If a one-sided interval is desired, the value K_1 is for the lower limit and K_2 is for use in determining an upper limit. The abscissa is given as "degrees of freedom for S." When the formula for S given in this section is used, degrees of freedom for S is equal to n. The values of γ are for two-sided intervals. For a one-sided interval the curve is selected as follows: If γ' is the desired confidence coefficient, use the curve for $\gamma = 2\gamma' - 1$.

d. Confidence intervals are calculated by substitution in the appropriate formula:

$$K_1 S \leq \sigma \leq K_2 S \quad \text{two-sided interval}$$

$$K_1 S \leq \sigma \quad \text{one-sided interval (lower limit)}$$

$$\sigma \leq K_2 S \quad \text{one-sided interval (upper limit)}$$

3.1.3. The Standard Deviation Of A Normal Population, Mean Unknown

An experiment is conducted to estimate the standard deviation of the population. The mean is unknown and will be estimated from the data.

μ = unknown population mean

σ = unknown population standard deviation

n = number of observations

X_i ($i = 1, 2, \dots, n$) = the observations

$$\bar{X} = \frac{1}{n} \sum X_i, \text{ the sample mean}$$

$$S = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}, \text{ the sample standard deviation}$$

$$\text{or } S = \sqrt{\frac{n \sum X_i^2 - (\sum X_i)^2}{n(n - 1)}}, \text{ computing formula}$$

a. The type of confidence interval appropriate for the desired statement is determined:

- 1) A two-sided interval if it is desired to place σ between two positive, finite limits.
- 2) A one-sided interval if it is desired to place a lower positive limit on σ , or to place an upper limit on σ

The confidence coefficient applies to any one of the three possible intervals but not to two or more simultaneously.

b. The desired confidence coefficient, γ , is selected. Values of γ available from the graphs are 0.80, 0.90, 0.95, and 0.99 for two-sided intervals and 0.90, 0.95, 0.975, and 0.995 for one-sided intervals.

c. Figure 3 contains a pair of curves for each value of γ . If a two-sided interval is desired, a value K_1 is read from the lower curve of the pair and K_2 is read from the upper curve. If a one-sided interval is desired, the value K_1 is for the lower limit and K_2 is for use in determining an upper limit. The abscissa is given as "degrees of freedom for S." When the formula for S given in this section is used, degrees of freedom for S is equal to $n - 1$. (Note: The formula in this section differs from the formula for S given in 3.4. When the population mean is unknown and estimated from the observations, the number of linearly independent variables in S is reduced by one. Hence degrees of freedom is one less than the sample size.) The values of γ are for two-sided intervals. For a one-sided interval the curve is selected as follows: If γ' is the desired confidence coefficient, use the curve for $\gamma = 2\gamma' - 1$.

d. Confidence intervals are calculated by substitution in the appropriate formula:

$$K_1 S \leq \sigma \leq K_2 S \text{ two-sided interval}$$

$$K_1 S \leq \sigma \quad \text{one-sided interval (lower limit)}$$

$$\sigma \leq K_2 S \text{ one-sided interval (upper limit)}$$

Illustration: The ability of a military vehicle to travel through soft soil may be judged by driving the vehicle along an appropriate course. The strength of the soil of the course may be inferred from cone penetrometer tests made immediately prior to the vehicle operation. The cone penetrometer result is expressed as a cone index. Uniformity of soil conditions over the test site is desirable. An estimate of variation in the course is to be inferred from the variation in cone index from location to location along the course. Measurements were made at 15 locations. From these a two-sided confidence interval for the standard deviation of cone index would be made. A 90 percent level was chosen. The cone indexes are: 25, 40, 42, 34, 33, 39, 32, 41, 34, 44, 31, 24, 48, 23 and 33.

σ = unknown population parameter, standard deviation of cone index

$n = 15$

$$\begin{aligned}
 S &= \sqrt{\frac{n \sum X^2 - (\sum X)^2}{n(n-1)}} \\
 &= \sqrt{\frac{15(19011) - (523)^2}{15(14)}} \\
 &= 7.4
 \end{aligned}$$

Referring to Figure 3, $\gamma = 0.90$ gives $K_1 = 0.77$ and $K_2 = 1.46$ for degrees of freedom $n - 1 = 14$.

$$K_1 S \leq \sigma \leq K_2 S$$

$$0.77(7.4) \leq \sigma \leq 1.46(7.4)$$

$$5.7 \leq \sigma \leq 10.8$$

3.1.4 Difference Between Means of Two Normal Populations, Standard Deviations Known, Sample Sizes Equal

An experiment is conducted to estimate the difference between the means of two normal populations, population A and population B. From prior experiments or for theoretical reasons the standard deviations of the populations are known. The sample sizes are equal, i.e., the number of observations from population A is the same as the number from population B. (This restriction on sample size is made in order to employ a simple graphical aid.)

μ_1 and μ_2 = unknown means of population A and B, respectively

σ_1 and σ_2 = known standard deviations

n = number of observations from each population

X_i ($i = 1, 2, \dots, n$) = observations from A

Y_i ($i = 1, 2, \dots, n$) = observations from B

$$\left. \begin{aligned}
 \bar{X} &= \frac{1}{n} \sum X_i \\
 \bar{Y} &= \frac{1}{n} \sum Y_i
 \end{aligned} \right\} \text{ sample means}$$

a. The type of confidence interval appropriate for the intended statement is selected.

- 1) A two-sided interval is the only appropriate interval if the sign of the difference $\mu_1 - \mu_2$ is of no interest.
- 2) Either the two-sided or a one-sided interval may be appropriate

if the difference of interest was selected before the experiment was run or the sign of the difference is a matter of consideration.

- 3) The two-sided interval places $\mu_1 - \mu_2$ between two finite limits, whereas a one-sided interval places either a lower or upper limit on $\mu_1 - \mu_2$.

b. The desired confidence coefficient, γ , is selected. Values of γ available from the graphs are 0.80, 0.90, 0.95, and 0.99 for two-sided intervals and 0.90, 0.95, 0.975, and 0.995 for one-sided intervals.

c. The curves of Figure 1 are used to obtain a factor K (ordinate) by selecting the appropriate curve and entering the graph with n as the abscissa. Curves are labeled with the confidence coefficients for two-sided intervals. For a one-sided interval the curve is selected as follows: If γ' is the desired confidence coefficient, use the curve for $\gamma = 2\gamma' - 1$.

d. Confidence intervals are calculated by substitution in the appropriate formula:

$$\text{Two-sided interval } (\bar{X} - \bar{Y}) - K \sqrt{\sigma_1^2 + \sigma_2^2} \leq \mu_1 - \mu_2 \leq (\bar{X} - \bar{Y}) + K \sqrt{\sigma_1^2 + \sigma_2^2}$$

$$\text{One-sided interval } (\bar{X} - \bar{Y}) - K \sqrt{\sigma_1^2 + \sigma_2^2} \leq \mu_1 - \mu_2$$

(lower limit)

$$\text{One-sided interval } \mu_1 - \mu_2 \leq (\bar{X} - \bar{Y}) + K \sqrt{\sigma_1^2 + \sigma_2^2}$$

(upper limit)

Illustration: Armor plate of two different aluminum alloys was tested to determine the difference between the alloys in resistance to penetration by cal .30 armor-piercing projectiles. Three sample plates of each alloy, all having the same areal density, were tested to find V-50 ballistic limits. Information from prior testing of similar material indicates that the standard deviation of ballistic limits would be 30 fps for either alloy when tested with this projectile. It is desired to estimate the difference between mean ballistic limits of the two alloys with 95 percent confidence limits. A two-sided interval is considered appropriate for the purpose of the test.

μ_1 = unknown mean ballistic limit of Alloy A

μ_2 = unknown mean ballistic limit of Alloy B

$\sigma_1 = \sigma_2 = 30$ fps

The test ballistic limits, in feet per second, are as follows:

<u>Alloy A</u>	<u>Alloy B</u>
1575	1559
1616	1502
1579	1511

$$\bar{X} = 4770/3 = 1590 \text{ (Alloy A)}$$

$$\bar{Y} = 4572/3 = 1524 \text{ (Alloy B)}$$

In Figure 1 the curve for $\gamma = 0.95$ is selected. The value $K = 1.13$ is obtained for $n = 3$. The two-sided interval is

$$(\bar{X} - \bar{Y}) - K \sqrt{\sigma_1^2 + \sigma_2^2} \leq \mu_1 - \mu_2 \leq (\bar{X} - \bar{Y}) + K \sqrt{\sigma_1^2 + \sigma_2^2}$$

$$(1590 - 1524) - 1.13 \sqrt{900 + 900} \leq \mu_1 - \mu_2 \leq (1590 - 1524) + 1.13 \sqrt{900 + 900}$$

$$66 - 1.13(42.4) \leq \mu_1 - \mu_2 \leq 66 + 1.13(42.4)$$

$$18 \leq \mu_1 - \mu_2 \leq 114$$

It is asserted with 95 percent confidence that the difference between population means of ballistic limits of these two aluminum alloys when tested with cal .30 armor-piercing projectiles is between 18 and 114 fps.

3.1.5 Difference Between Means of Two Normal Populations, Standard Deviations Unknown But Equal, Sample Sizes Equal

An experiment is conducted to estimate the difference between the means of two normal populations, populations A and B. Standard deviations are unknown but can be assumed to be the same for both populations. The sample sizes are equal, i.e., the number of observations from population A is the same as the number from population B. (This restriction on sample size is made in order to employ a simple graphical aid.)

μ_1 and μ_2 = unknown means of populations A and B, respectively

n = number of observations from each population

X_i ($i = 1, 2, \dots, n$) = observations from A

Y_i ($i = 1, 2, \dots, n$) = observations from B

$$\left. \begin{aligned} \bar{X} &= \frac{1}{n} \sum X_i \\ \bar{Y} &= \frac{1}{n} \sum Y_i \end{aligned} \right\} \text{ sample means}$$

$$s_1^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1}, \text{ sample variance from A}$$

$$\text{or } S_1^2 = \frac{n \sum X_1^2 - (\sum X_1)^2}{n(n-1)}, \text{ computing formula}$$

$$S_2^2 = \frac{\sum (Y_1 - \bar{Y})^2}{n-1}, \text{ sample variance from B}$$

$$S_p = \sqrt{\frac{S_1^2 + S_2^2}{2}}$$

a. The type of confidence interval appropriate for the intended statement is selected.

- 1) A two-sided interval is the only appropriate interval if the sign of the difference $\mu_1 - \mu_2$ is of no interest.
- 2) Either the two-sided or a one-sided interval may be appropriate if the difference of interest was selected before the experiment was run or the sign of the difference is a matter of consideration.
- 3) The two-sided interval places $\mu_1 - \mu_2$ between two finite limits, whereas a one-sided interval places a lower or upper limit on $\mu_1 - \mu_2$.

b. The desired confidence coefficient, γ , is selected. Values of γ available from the graphs are 0.80, 0.90, 0.95, and 0.99 for two-sided intervals and 0.90, 0.95, 0.975, and 0.995 for one-sided intervals.

c. The curves of Figure 4 are used to obtain a factor K (ordinate) by selecting the appropriate curve and entering the graph with n as the abscissa. Curves are labeled with the confidence coefficients for two-sided intervals. For a one-sided interval the curve is selected as follows: If γ' is the desired confidence coefficient, use the curve for $\gamma = 2\gamma' - 1$.

d. Confidence intervals are calculated by substitution in the appropriate formula:

$$(\bar{X} - \bar{Y}) - KS_p \leq \mu_1 - \mu_2 \leq (\bar{X} - \bar{Y}) + KS_p \text{ two-sided interval}$$

$$(\bar{X} - \bar{Y}) - KS_p \leq \mu_1 - \mu_2 \text{ one-sided interval (lower limit)}$$

$$\mu_1 - \mu_2 \leq (\bar{X} - \bar{Y}) + KS_p \text{ one-sided interval (upper limit)}$$

Illustration: Preliminary information indicated that the addition of a certain wear reducing agent to the standard propelling charge for 175-mm ammunition produced an increase in initial blast overpressure. An estimate of the increase was desired. Eight rounds with the standard charge and eight with the agent added to the standard charge were fired. Blast overpressure occurring at a selected position near the weapon was recorded for each round. (For simplicity of the illustration only one location is considered.) It is desired to place 95 percent confidence limits on the difference between means of blast overpressure. The standard deviations are unknown but believed to be equal.

The blast overpressures (psi) are as follows:

<u>Standard Charge With Additive</u>	<u>Standard Charge</u>
2.26	1.17
2.16	1.05
2.07	1.23
2.11	1.03
1.98	1.27
2.36	1.21
2.16	1.31
2.19	1.12

μ_1 = unknown mean overpressure, charge with additive

μ_2 = unknown mean overpressure, standard charge

$n = 8$

$$\bar{X} = \frac{17.29}{8} = 2.16 \text{ (with additive)}$$

$$\bar{Y} = \frac{9.39}{8} = 1.17 \text{ (standard)}$$

$$s_1^2 = \frac{8(37.4619) - (17.29)^2}{8(7)} = 0.0134$$

$$s_2^2 = \frac{8(11.0927) - (9.39)^2}{8(7)} = 0.0102$$

$$s_p = \sqrt{\frac{0.0134 + 0.0102}{2}} = 0.11$$

Reference to Figure 4 for $n = 8$ and $\gamma = 0.95$ gives $K = 1.07$. The two-sided confidence interval is

$$(2.16 - 1.17) - 1.07(0.11) \leq \mu_1 - \mu_2 \leq (2.16 - 1.17) + 1.07(0.11)$$

$$0.87 \leq \mu_1 - \mu_2 \leq 1.11$$

It is asserted with 95 percent confidence that the average increase in blast overpressure, at the specific location, resulting from the use of this wear-reducing agent is between 0.87 and 1.11 psi.

3.1.6 Binomial P

When the result of a single trial in an experiment is one of two mutually exclusive categories such as "success" or "failure", and the probability of failure in a single trial is constant from trial to trial, the binomial distribution is

applicable. This distribution contains one parameter, P ($0 \leq P \leq 1$), which is the probability of failure in a single trial. If n is the number of trials conducted, the probabilities of the number of failures ($P, 1, 2, \dots, n$) can be calculated in terms of n and P . (In this discussion P is probability of failure, but it could be probability of success by redefinition of categories).

Since P is a population parameter it is usually unknown. Interval estimates of P can be obtained by observing the number of successes in a sample of n trials. The calculations, however, are not simple for the general case. Therefore only one special situation is treated in this pamphlet. This is the case in which P is expected to be close to zero. The extremes of the possible range of P are of most frequent interest in testing materiel. Often P represents the probability of failure of an item when it is employed in a specified environment.

An experiment consisting of several trials is performed in order to find an upper confidence limit on P .

n = number of trials

K = number of failures observed

P = unknown population parameter, probability of failure
in a single trial

P_u = upper confidence limit for P

- a. The figure corresponding to K (0, 1, 2, 3, or 4) is selected from among Figures 5 through 9.
- b. The desired confidence coefficient is selected: $\gamma = 0.90, 0.95$, or 0.99 .
- c. The graph is entered with n as abscissa and the value P_u (ordinate) is read from the curve for the selected value of γ .
- d. The statement is made that $P \leq P_u$ with 100γ percent confidence.

Illustration: A sample of 48 projectiles, HE, M437 selected from one production lot were subjected to a standard laboratory vibration schedule to determine what proportion of that lot could be expected to suffer damage in that environment. (Production is assumed to be large relative to the sample size of 48.) After the vibration test was completed, inspection of the projectiles revealed that none of the 48 was damaged.

P = unknown binomial parameter representing probability of
an individual projectile's being damaged

n = 48

K = 0

A 95 percent upper confidence limit on P is desired. The appropriate graph for $K = 0$ and $\gamma = 0.95$ is found in Figure 5. The graph is entered with

MTP 3-1-002
25 January 1967

$n = 48$ as the abscissa, and a value P_u (ordinate) is read from the curve for $\gamma = 0.95$ as 0.060. The interval is $P \leq 0.060$. It is asserted with 95 percent confidence that the probability of a projectile of this lot's being damaged in this vibrational environment is no greater than 0.060.

REFERENCES

1. AMCP 706-110 through -114, Engineering Design Handbook, Experimental Statistics, Secs. I through V, respectively, July 1963.
2. U. S. Department of Commerce, Experimental Statistics: Handbook 91, 1 August 1963.
3. U. S. Army Test and Evaluation Command Materiel Test Procedures, MTP 3-2-828, Statistical Aid.
4. Simon, L. E., An Engineer Manual of Statistical Methods, John Wiley and Sons, Inc.

APPENDIX A

GRAPHS FOR CALCULATING CONFIDENCE INTERVALS

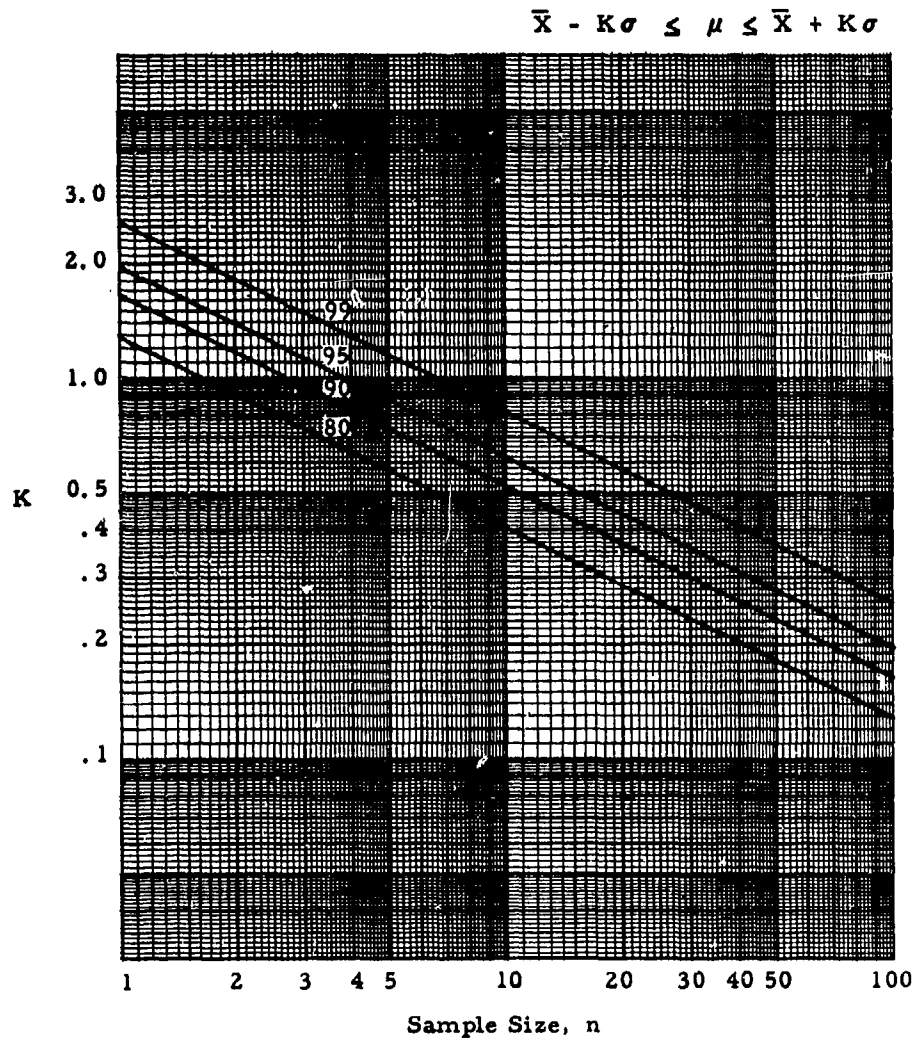


Figure 1. Factors K, for Estimating the Mean, μ , of a Normal Population - Population Standard Deviation, σ , Known. Confidence Coefficients (Two-Sided Interval) $\gamma = 0.80, 0.90, 0.95$, and 0.99 .

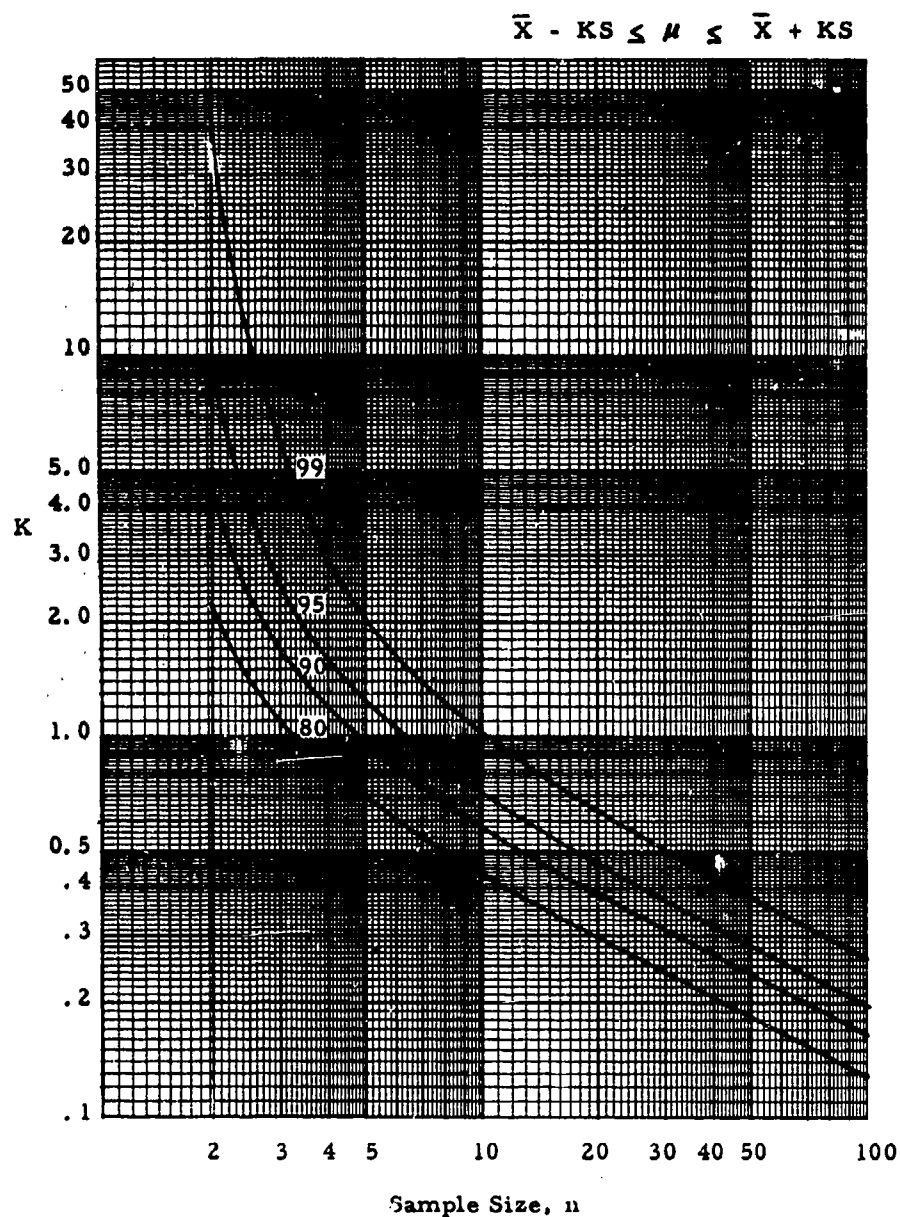


Figure 2. Factors, K, for Estimating the Mean, μ , of a Normal Population - Population Standard Deviation, σ , Unknown and Estimated by Sample Standard Deviation S. Confidence Coefficients (Two-Sided Interval) $\gamma = 0.80, 0.90, 0.95$, and 0.99 .

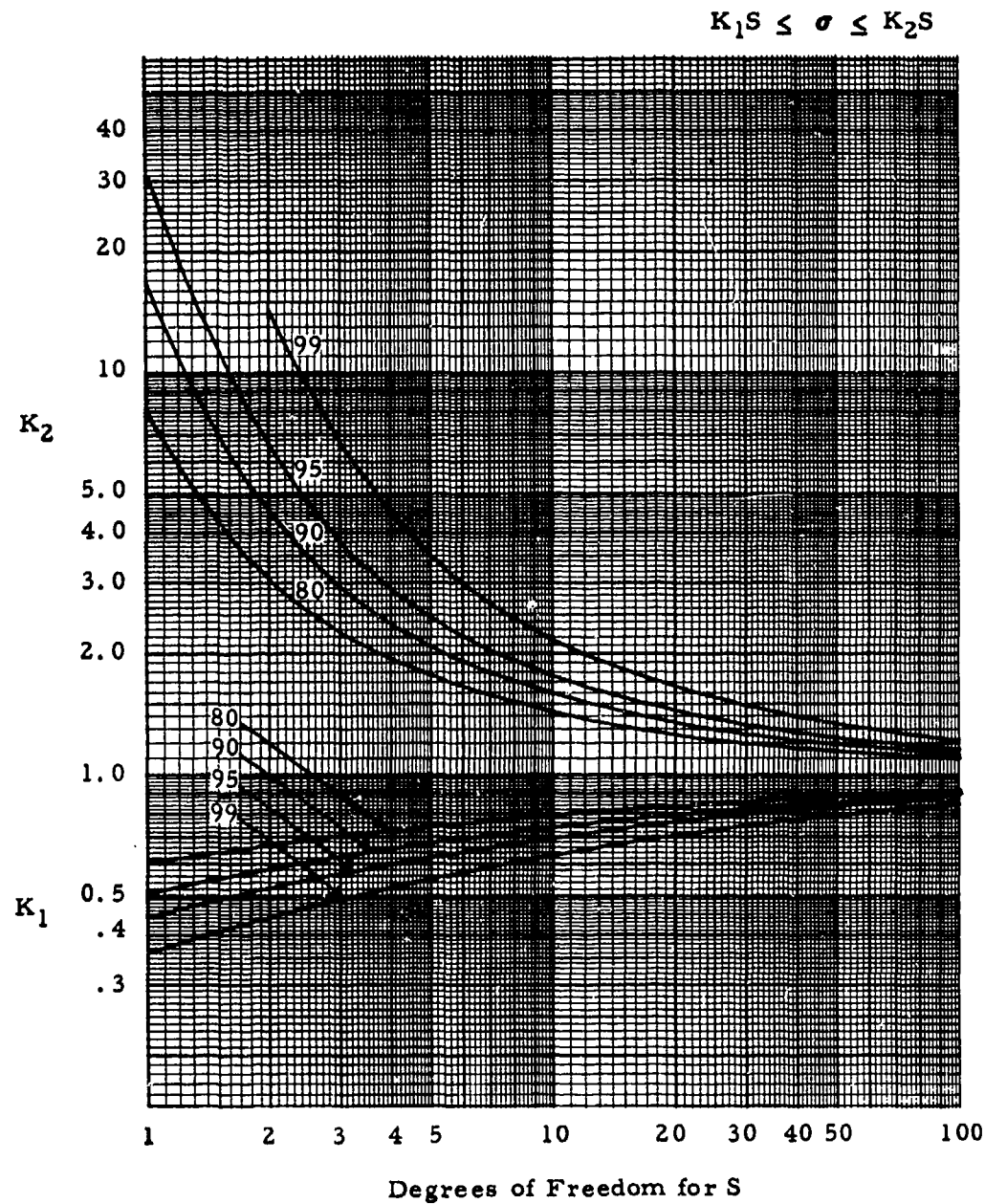


Figure 3. Factors K_1 and K_2 , for Estimating the Standard Deviation, σ . Confidence Coefficients (Two-Sided Interval) $\gamma = 0.80, 0.90, 0.95$, and 0.99 .

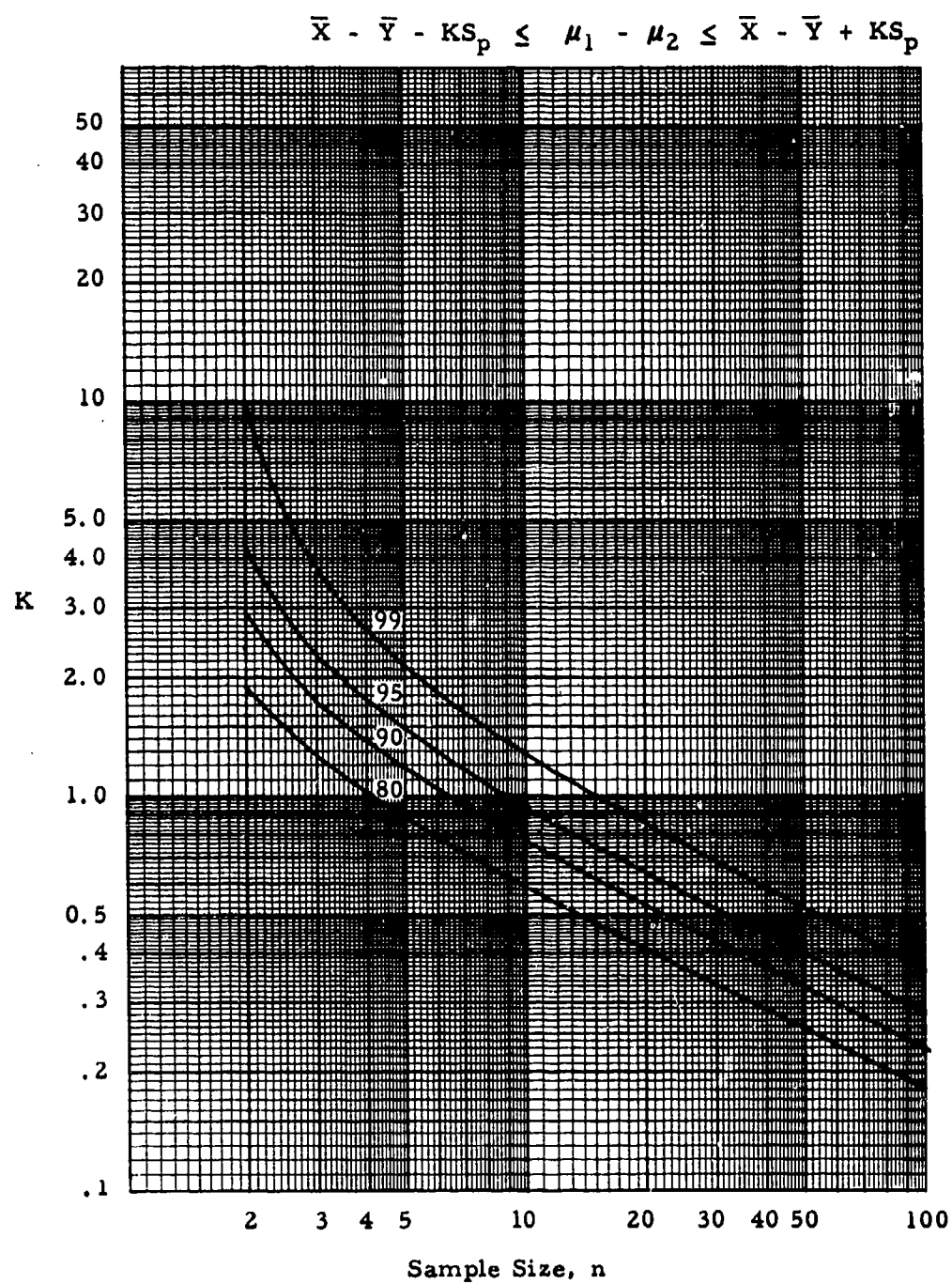


Figure 4. Factors, K, for Estimating the Difference, $\mu_1 - \mu_2$, Between Means of Two Normal Populations - Standard Deviations Unknown But Equal; Equal Sample Sizes. Confidence Coefficients (Two-Sided Interval) $\gamma = 0.80, 0.90$, and 0.99 .

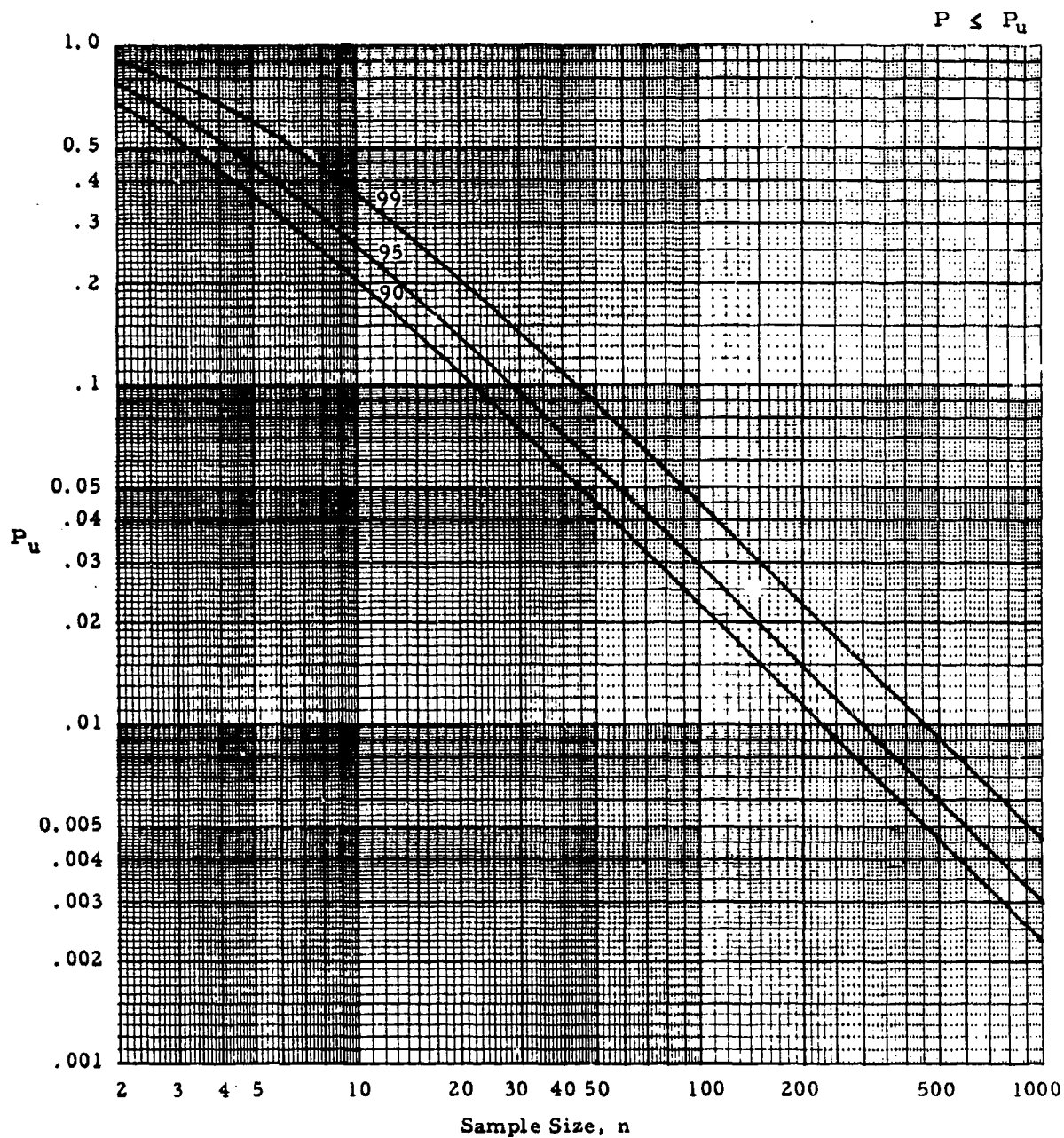


Figure 5. Upper Confidence Limit, P_u , of Binomial P (P = Probability of Failure in One Trial) When No Failure is Observed in n Trials. Confidence Coefficients $\gamma = 0.90, 0.95$, and 0.99 .

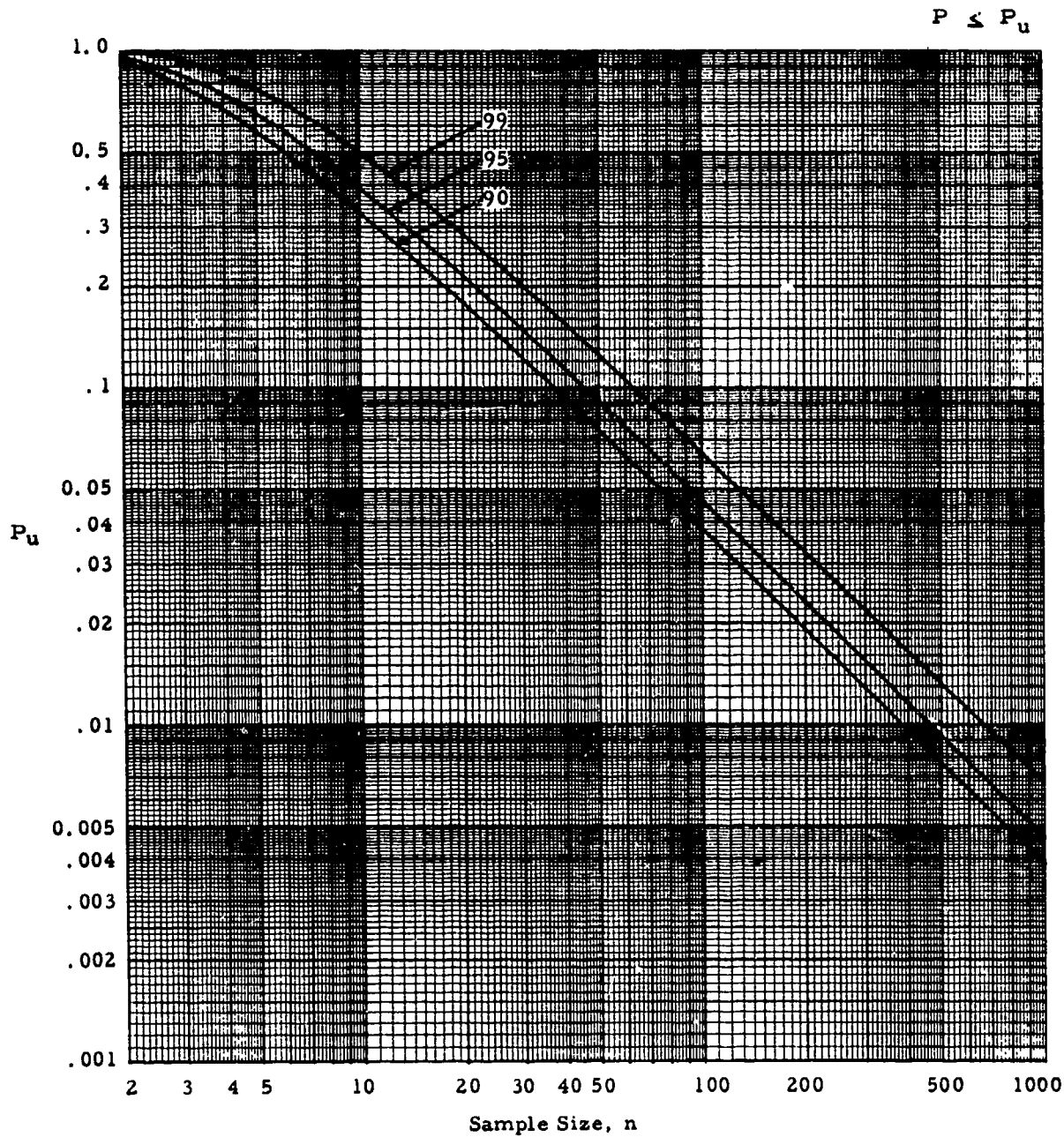


Figure 6. Upper Confidence Limit, P_u , of Binomial P (P = Probability of Failure in One Trial) When One Failure is observed in n Trials. Confidence Coefficients $\gamma = 0.90, 0.95$, and 0.99 .

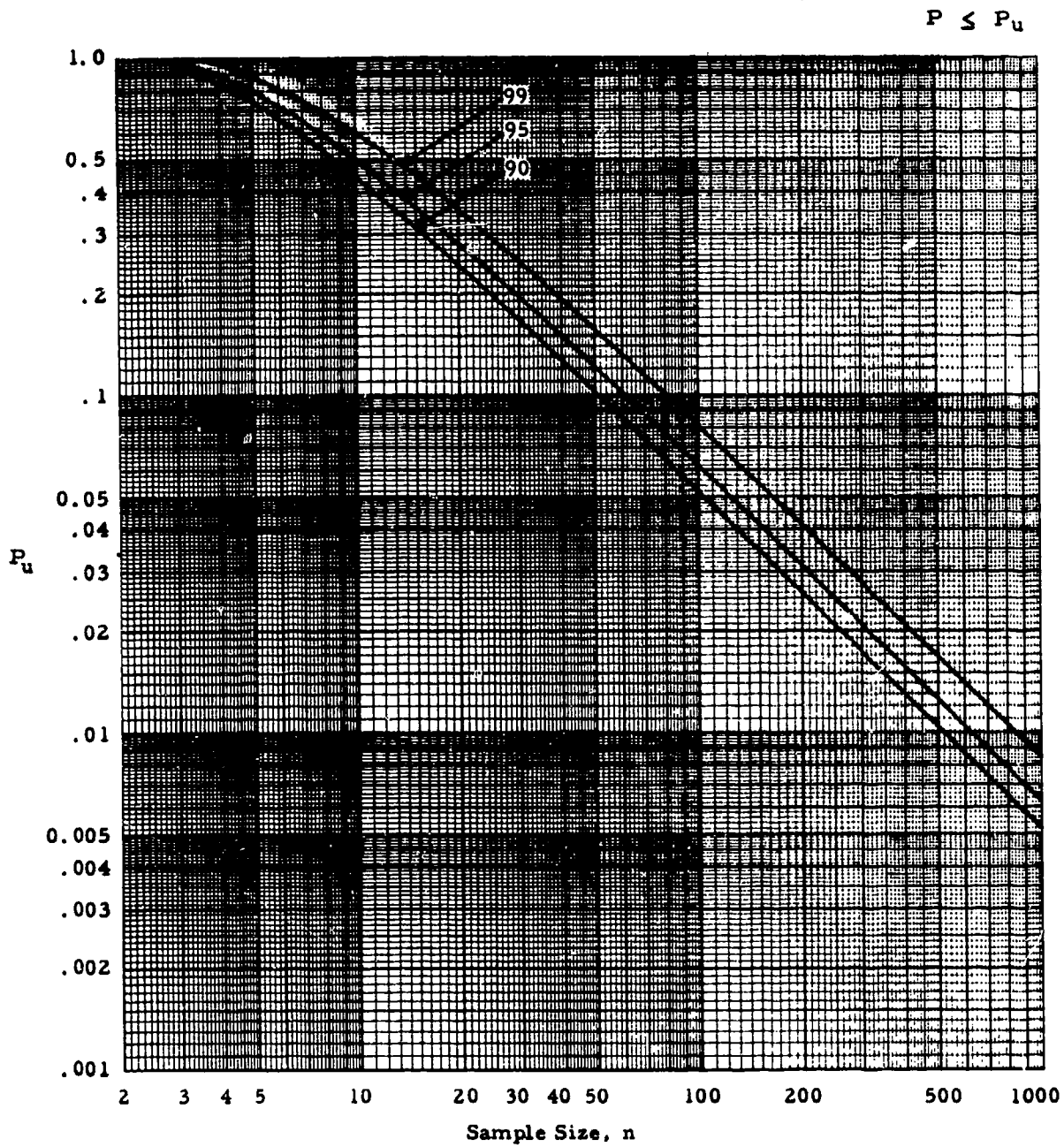


Figure 7. Upper Confidence Limit, P_u , of Binomial P (P = Probability of Failure in One Trial) When Two Failures are Observed in n Trials. Confidence Coefficients $\gamma = 0.90, 0.95$, and 0.99

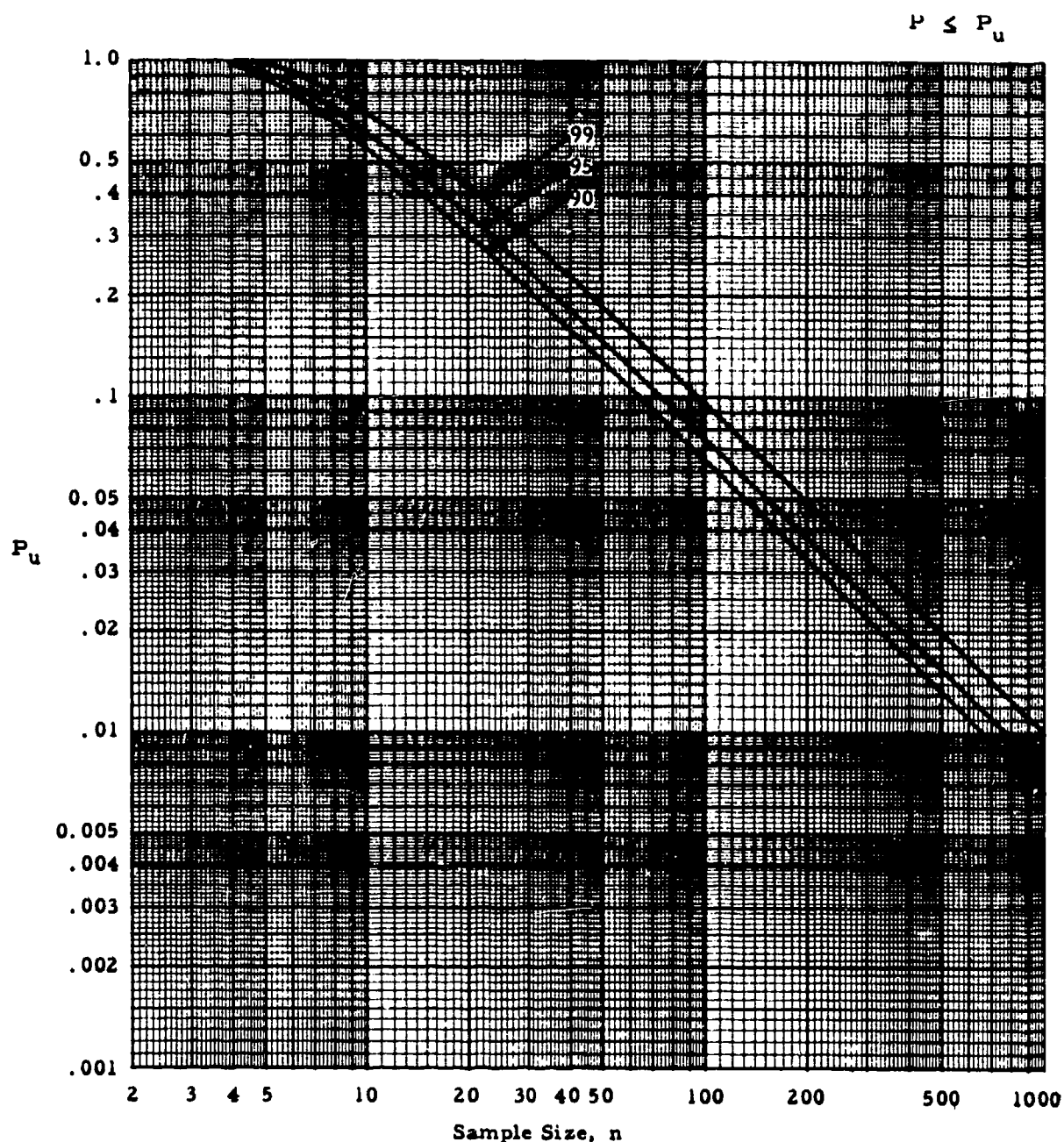


Figure 8. Upper Confidence Limit, P_u , of Binomial P (P = Probability of Failure in One Trial) When Three Failures are Observed in n Trials. Confidence Coefficients $\gamma = 0.90, 0.95$, and 0.99 .

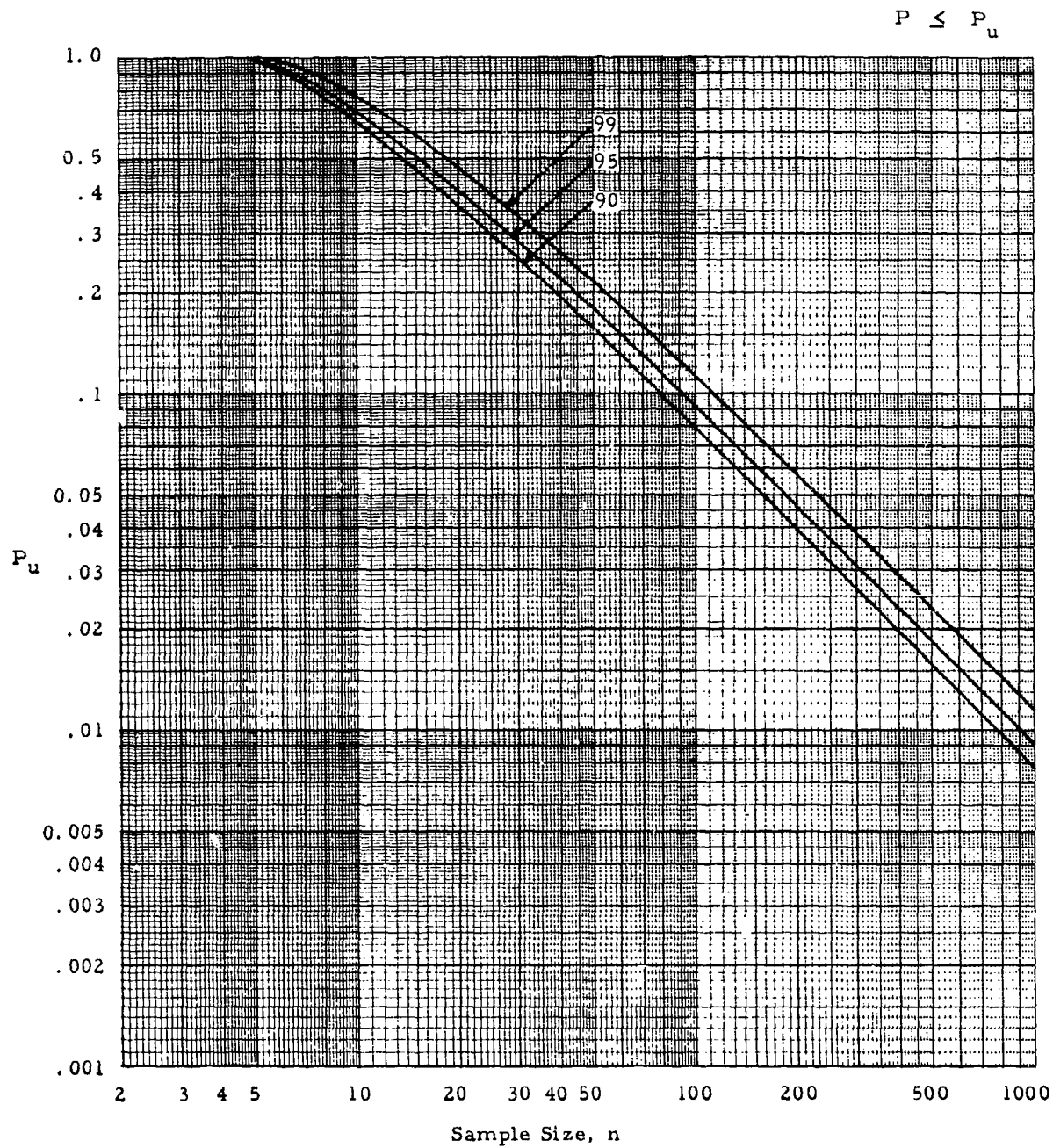


Figure 9. Upper Confidence Limit, P_u , of Binomial P (P = Probability of Failure in One Trial) When Four Failures are Observed in n Trials. Confidence Coefficients $\gamma = 0.90, 0.95$, and 0.99 .

ADDITIONAL FOR

OFSTI WHITE SECTION ☒

DC BUFF SECTION ☐

ANNOUNCED ☐

LOCATION

per form 50

DISTRIBUTION/AVAILABILITY CODES

DIST.	AVAIL. and/or SPECIAL
<i>[scribble]</i>	<i>1</i>